

App. No. 10/042,880
Art Unit: 2654

Docket No. 2001-0410

In the Specification:

Kindly amend paragraph 38 of the specification as follows:

[38] If the look-ahead delay is to be minimized, the most efficient method to locate the TSMS is a cross-correlation measure that [[it]] can be obtained directly from WSOLA synthesis of the previous frame. However, considerable performance improvements (fewer boundary errors and/or distortions in the modified speech) are realized when the TSMS duration is known before its modification begins; hence the reduced complexity advantages cannot be capitalized upon. Both the correlation and energy measures can identify long duration high-energy speech sections of the signal that correspond to voiced portions to be modified. The short-time energy, E_n , of the signal $x(l)$ centered at time $t=n$, is calculated as

$$E_n = \sqrt{\frac{1}{N+1} \sum_{m=-N/2}^{N/2} x^2(n+m)} \quad (4)$$

where the window length $N=20$ ms. However, time-domain measures have difficulty discriminating two syllables in a continuous voiced section. TSMS detection is more reliably accomplished using a measure that detects abrupt changes in frequency-domain characteristics, such as the known spectral feature transition rate (SFTR). The SFTR is calculated as the gradient, at time n , between the Line Spectral Frequencies (LSFs), y_l , within the interval $[n \pm M]$. This is given by the equation:

$$SFTR = \sum_{l=1}^P (g_n^l)^2 \quad (5)$$

where, the gradient of the l^{th} LSF, is

$$g_n^l = \frac{\sum_{m=-M}^M m y_l(n+m)}{\sum_{m=-M}^M m^2}, \quad l = 1, \dots, P \quad (6)$$

and P , the order of prediction, is 10. LSFs are calculated every 10ms using a window of 30ms. The SFTR can then be mapped to a value in the range $[0, 1]$, by the function:

App. No. 10/042,880
Art Unit: 2654

Docket No. 2001-0410

$$C_n = \frac{2}{1 + e^{-\beta(SFTR)}} - 1 \quad (7)$$

where, the variable β is set to 20.